Technical and Editorial Administration of a World-Wide-Web Site During a Period of Rapid Growth: The OncoLink® Experience

Ivor Benjamin, M.D.*, Thomas J. Dilling, B.A.†, Keith C. Campbell, M.S.E.+, Ammar Maraqa+,
Bill Liang, M.S.E.+, Robbie Medbery, M.D.†, Joel W. Goldwein, M.D.†
Departments of *Obstetrics & Gynecology and †Radiation Oncology at the School of Medicine
University of Pennsylvania Cancer Center
+University of Pennsylvania School of Engineering and Applied Science
Philadelphia, Pennsylvania

OncoLink® is a cancer information resource on the World-Wide-Web (WWW) that provides a wide variety of information for cancer patients and healthcare providers. Since its introduction in March, 1994 it has enjoyed success as demonstrated by an over 31-fold increase in usage as of February, 1996. Current utilization exceeds 1.1 million accesses per month. The content of OncoLink has also expanded greatly, with new items being added daily. In addition, OncoLink has been the recipient of numerous awards from a variety of agencies and organizations.

During this period of rapid growth, the complexity of managing and maintaining OncoLink has likewise increased. This work may be divided into three categories: content editing, technical (or production) editing, and web site maintenance. Consequently, we have developed numerous administrative procedures to handle this workload. After implementing these new administrative strategies, we were able to greatly reduce the need for face-to-face meetings of our Editorial and Production Staffs.

This paper describes our experience with developing efficient strategies for managing the daily operation of OncoLink during a period of rapid growth.

INTRODUCTION

The World-Wide-Web (WWW) is becoming an increasingly popular venue for the dissemination of medical information. (1, 2) Presently, there are thousands of medically oriented Web sites cataloged in the Medical Matrix, (3) Yahoo! (4) and Lycos, Inc. (5) indices. These medical web resources vary greatly in depth and quality of content.

We have previously reported on the early development of OncoLink® (http://www.oncolink.upenn.edu/) as a multimedia oncology information resource on the Internet.(6) Since that report, OncoLink's content has tripled. Popularity, as measured by monthly usage, had increased by over 31-fold with current demand exceeding 1.1 million accesses per month. In addition to being the recipient of numerous awards, (7)

OncoLink has been mentioned on several occasions by the press, news media and professional medical publications. (8-11) The University of Pennsylvania Cancer Center (UPCC), one of 26 centers in the United States designated by the National Cancer Institute as a Comprehensive Cancer Center, is responsible for the development and oversight of OncoLink.

In order to manage this increased content and traffic, we have developed several administrative procedures for the maintenance of OncoLink. These may be divided into three categories: 1. evaluation of new content from an editorial perspective, 2. processing of accepted material for inclusion into the menu structure and hierarchy of related links and 3. Web site maintenance (e.g. log file analysis, link expirations, etc.).

When developing these administrative operating procedures, we had the following goals: 1. maximize automation of communication within the OncoLink Editorial Staff by leveraging the WWW tools available in HTML v2.0, 2. devise efficient, unambiguous communication that would limit the need for face-to-face meetings by the Editorial Staff, 3. allow for future expansion, and 4. maintain portability (i.e. not be platform dependent).

METHODS

The administration of OncoLink is conducted by an Editorial Staff, which is composed of an Editorial Board and Production Editors. The Editorial Board is composed of Board certified and/or eligible oncologists on the faculty of UPCC who are responsible for the content of OncoLink. An Editorial Assistant works with the Editorial Board and is primarily responsible for copyright clearances and document handling. The Production Editors are students at the University of Pennsylvania School of Engineering and Applied Science and others with strong UNIX, Perl, HTML and database management experience. The Production Editors are responsible for HTML coding, maintaining the Linklog and Documents databases, developing administrative WWW forms,

398

maintaining the Web server, and responding to user inquiries of a technical nature.

The Production Editors and Staff spend a total of 20-40 hours weekly on OncoLink. Additional time is spent by one team member who is working on a grant-supported project which will be incorporated into OncoLink at a later date. The two chief editors spend between 2-3 hours daily dividing their time equally between responding to e-mail and acquiring and editing content. The Editorial Board and the Production Editors are jointly responsible for the look and feel of the WWW pages. Several legal, administrative and non-administrative members of the University of Pennsylvania Cancer Center serve in an advisory capacity. In total, they commit approximately 15 hours monthly on the projects.

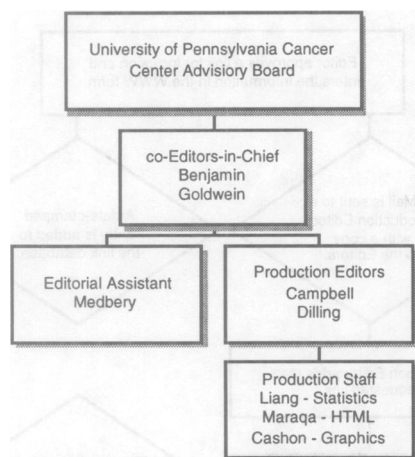The organizational structure of OncoLink's Editorial Staff is shown in Figure 1.

Figure 1. Organizational structure of OncoLink's Editorial Staff

Automated Communications

In order to coordinate workflow and facilitate communications between the Editorial Board and Production Editors, we have created several administrative WWW forms which access databases of information and enable us to track workflow. We also rely extensively on automated e-mail, the content of which is the data entered into the administrative WWW forms. We have minimized the need to manually write e-mail messages for communication of workflow issues, increasing efficiency and minimizing the necessity for face-to-face meetings of the Editorial Staff. Ordinarily, the members of the Editorial Staff are located in distant buildings on our campus and on several occasions we have been separated by thousands of miles for weeks or months at a time. While on the road, these Editors have continued their duties via a laptop

computer by establishing dial-up connections into a local point of presence via an Internet service provider.

In general, our workflow falls into two distinct categories: 1. evaluation and approval of new links and proprietary content from an editorial perspective by the Editorial Board and 2. the addition of new links to external web sites and addition of new proprietary content to our server by the Production Editors. Several WWW forms and two databases (Linklog and document database) have been created to manage these processes.

Editorial Process

The Editorial Board is responsible for the content of OncoLink. All material is reviewed by members of the Editorial Board with the expert assistance of reviewers from within the University of Pennsylvania Cancer Center and reviewers from other institutions. This editorial process is as follows:

1) An author submits a document, which is read by an Editor-in-Chief and receives preliminary approval for posting from the Editorial Board.

2) The Editorial Board then assigns the document to reviewers with appropriate expertise.

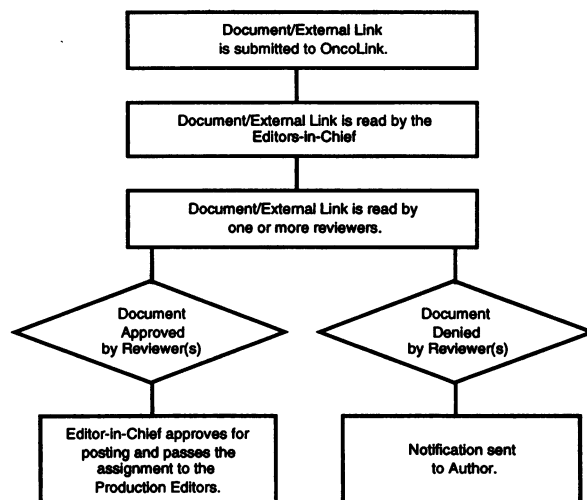The workflow for the editorial process is shown in Figure 2.

Figure 2. Flowchart for the editorial process.

Adding new links: The Linklog database

All new links to other web sites are added to OncoLink web pages upon the approval of the Editorial Board. Our administrative WWW forms are located in a "private" directory on OncoLink that is unpublished and to which access is restricted for security reasons.

After approval of an appropriate link, the Editor calls the WWW form created for the addition of new links. This form is shown in Figure 3 and has fields for the following information:

1) name (descriptor) for the link;
2) URL of the link;
3) editorial annotation of the link;
4) location(s) for the link within the OncoLink hierarchy
5) name of the editor who is requesting the link
6) any additional comments for the production staff

Figure 3. The WWW form for adding a new link.

Submission of this data activates a Perl script which performs two tasks: (1) An automated e-mail message is sent to the Production Editor who specifically handles addition of new links. A copy of the message is sent to the members of the Editorial Board, all of whom maintain separate e-mail accounts that share a single e-mail alias. (2) The information from the WWW form is appended to an HTML file, thus creating a record in the Linklog database. The entry is also date-stamped. A sample entry from this database follows:

```
<DL>
```

```
<DT><A
HREF="http://www.westvirginia.com/wv002/nhl
home.htm">JL  Sigley:Non-Hodgkin's  Lymphoma
Web Page</A>
<DD>EXTERNAL Link Request generated on: Sun
Mar 24 10:34:54 1996 by JOEL
<DL><DT>Desired location(s) for the link:
<DD>
<DD>What's New
Diseases - Lymphoma
Global Resources - Other
<DT>Additional annotation:
<DD>This resource is dedicated specifically
to persons with Non-Hodgkin's Lymphoma. It
contains many useful WWW links as well as
personal stories by lymphoma patients.
</DL>
</DL>
```

Having received this assignment from the Editorial Board, the Production Editor adds the link where required, following the instructions provided. An overview of the entire process for adding a new link is outlined in Figure 4.
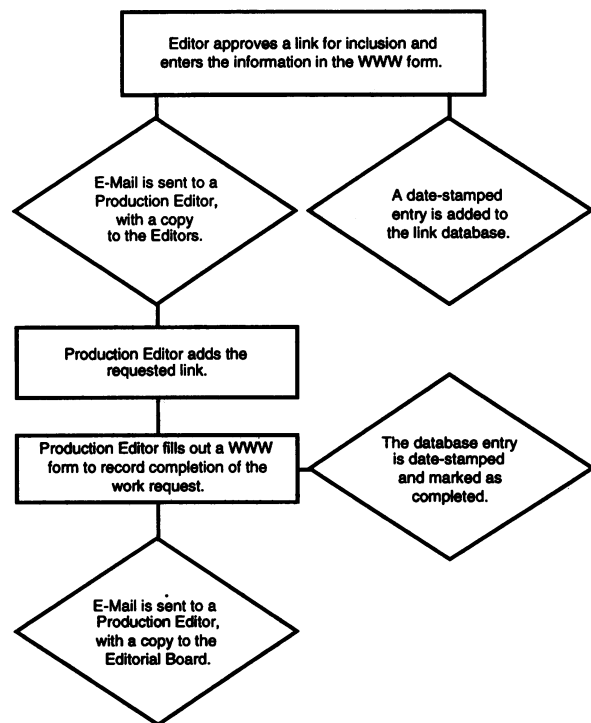
Figure 4. Flowchart for processing of a newly approved link.

After having added the link, the Production Editor completes another WWW form which requires entry into two fields: 1) URL of the link added and 2) name of the individual who placed the link on-line (i.e. his/her name). Submission of this form activates another Perl script which 1) notifies the Editorial Board via e-mail that the link has been added and sends a copy of this message to the Production Editor

originally assigned the task; and 2) revises the HTML file accordingly, adding a line which states who added the link, and at what time, as outlined below:

```
<DL>
<DT><A
HREF="http://www.westvirginia.com/wv002/nhl
home.htm">JL
Sigley:Non-Hodgkin's Lymphoma Web Page</A>
<DD>EXTERNAL Link Request generated on: Sun
Mar 24 10:34:54 1996 by JOEL
<DD>Link added by AMMAR on: Mon Mar 26
13:53:27 1996
<DL><DT>Desired location(s) for the link:
<DD>
<DD>What's New
Diseases - Lymphoma
Global Resources - Other
<DT>Additional annotation:
<DD>This resource is dedicated specifically
to persons with Non-Hodgkin's Lymphoma. It
contains many useful WWW links as well as
personal stories by lymphoma patients.
</DL></DL>
```

If any of the Editors wish to track the work assignments, they can easily scan the HTML file and note progress. When the file reaches a significant length, the completed entries are purged into a secondary file, for future reference as needed. This is accomplished via a Perl script which the Production Editors periodically call from the server's command line.

Adding new documents: Document database
The addition of proprietary content to OncoLink follows an overall scheme that is similar to adding external links though management of the information is more complex. In many cases, there are a series of communications must occur between the author and the Editorial Staff before the document is finally approved for posting on-line. The Document database is a textfile which records summary data about the document and status of the document in the approvals process. This database is manipulated by the OncoLink staff through use of a Perl script activated by use of various WWW forms in our private administrative directory.

The workflow is as follows:

1) Following preliminary approval from the Editorial Board, the Editorial Assistant enters summary data into the database, via a WWW form. She then sends a copyright transfer form to the author which must be returned after signature.

2) Upon receipt of the copyright form from the author, the Editorial Assistant records this in the database and assigns the material to a Production Editor for HTML coding.

3) When the Production Editor has placed the material on-line, he fills out another form with the final URL of the document. This information is added to the summary data for the Document database entry.

Log analysis and access statistics
The Production Editors utilize a readily-available shareware WWW log analysis package called wwwstat to create monthly statistical reports regarding utilization of the server.(12) Information is gathered as to the total number of accesses in a given month, as well as the number of unique IP addresses which accessed the server. This information is summarized at URL: http://www.oncolink.upenn.edu/stats/. We also gather information for our own use regarding the types of WWW browsers our users employ, as well as the external WWW pages from which they access our site. This information is summarized in another report which resides in our private administrative directory.

RESULTS

Utilization and access statistics
In its first month, OncoLink served approximately 35,000 files to users from 1,600 unique IP addresses world-wide. In February 1996, OncoLink was accessed over 1.1 million times from 58,800 IP addresses. This represents a growth of over 3,100% in a two year period. During this time, a total of three server upgrades have been required to support the heavy utilization along with 3 upgrades of server software. Currently, we are utilizing NCSA's HTTPd version 1.5 (13) on a DEC Alpha 2100 computer (Digital Equipment Corporation) (14) running under OSF/1. The OncoLink server has been "down" a total of 4 times in the last year. Only one incident exceeded 2 hours. In only 2 of these 4 cases, was the downtime unexpected. In 2 of 4 cases where downtime was scheduled, a standard operating procedure for prior announcement of downtime was implemented.

Benefits from automation
Organization of our operating procedures as cited above has been the most important outcome of our project. As a consequence of these procedures, each of the editors processes an average of approximately 15 e-mail messages daily which are related to OncoLink editorial activity. Most of these come from members of our Production Editors and Staff, and most require response. The Editorial Staff communicates via e-mail on a daily basis, including weekends and holidays. Coverage for personnel who are out of contact is routinely arranged and automated e-mail is redirected during these absences.

As a consequence of implementing the automated communications between Editorial and Production

Staffs, the face-to-face meeting time has been reduced by 75%. Prior to implementation of these strategies, the Editors-in-Chief held face-to-face meetings for 1-2 hours per week (up to 8 hours per month). Presently, with the automation schemes outlined in this paper, the Editorial Staff holds a single 2 hour monthly face-to-face meeting. Furthermore, the Editorial and Production Staffs are adding a large body of new material to OncoLink than was possible under the previous administration scheme. Formal meetings with the UPCC Advisory Board are held approximately 2-3 times per year.

Difficulties related to growth

On average, an additional 12-24 external messages from OncoLink users are received daily. These are answered by members of the Editorial Board. As popularity of our resource has grown, however, the volume of e-mail received has increased proportionally. Many of these letters are requests for specific medical advice. As a policy, we do not render medical consultations via the Internet. This is stated in the OncoLink disclaimer on the home page. In order to minimize the volume of e-mail sent to us, we have instituted several measures to point our readers to appropriate resources. The WWW form provided for users to send us e-mail includes several suggestions and pointers to assist them in finding the information that they are seeking. Since many of the e-mail messages which we receive are patients seeking explicit medical consultation, we restate our policy of not providing specific medical consultation via the Internet on the e-mail form.

To provide users with additional information about the purpose and intended usage of OncoLink, we have created an "About OncoLink" page accessible from the home page. On the "About OncoLink" page, we have provided information about the content of OncoLink and a RealAudio based synchronized multimedia "Guided Tour" of the site.

As mentioned elsewhere, content hosted on our server has tripled in two years. Our directory structure has also grown proportionally. The Editorial staff has found that much of the discussion at the monthly meetings has involved ongoing redesign of the WWW pages, to better accommodate the increasing volume of information, without resulting in diminishing the user's ability to find desired information quickly and easily.

DISCUSSION

In this paper we have described efficient strategies for managing a medically-oriented web site that is updated daily and incurs significant traffic. As a consequence of implementation of these automated strate-

gies, significant reductions in face-to-face meeting times were achieved while, in our opinion, maintaining the editorial quality of the content. We believe an automated process that leverages the power of the WWW and other Internet services is extremely useful in administering a medically-oriented web site in an academic institution. Since the members of the Editorial Board are busy oncologists with significant clinical responsibilities, automated communication is vital for maintaining a constant flow of critically reviewed, new material onto OncoLink's pages.

References

1. Glowniak JV. Medical resources on the Internet. Ann Intern Med 1995;123(2):123-31.
2. Goldwein JW, Benjamin I. Internet-based medical information: Time to take charge (Editorial). Ann Int Med 1995;123(2):152-3.
3. Medical Matrix. http://www.kumc.edu:80/mmatrix/ 1996;.
4. Yahoo! http://www.yahoo.com/Health/Medicine/ 1996;.
5. Lycos, Inc. http://www.lycos.com/ 1996;.
6. Buhle EL, Jr., Goldwein JW, Benjamin I. OncoLink: a multimedia oncology information resource on the Internet. Proc Annu Symp Comput Appl Med Care 1994;(1):103-7.
7. OncoLink Awards Page. http://www.oncolink.upenn.edu/stuff/oncolink/awards.html .
8. Jaroff L. The man's cancer. Time 1996 April 1, 1996:58-65.
9. Larkin M. Oncology on the Internet: Online know-how now a necessity for clinicians. Oncology Times 1995;17(12):10-17.
10. Winslow R. More doctors are adding on-line tools to their kits. Wall Street Journal 1994 Oct. 7, 1994:B1.
11. Gibson C. ABC Television's Good Morning America: To your health. 1995 November 21, 1995:
12. WWWStat. http://www.ics.uci.edu/WebSoft/wwwstat/ .
13. NCSA's HTTPd 1.5. http://hoohoo.ncsa.uiuc.edu/docs/Overview.html 1996;.
14. DEC Alpha 2100. http://www.digital.com:80/.i/info/alphaserver/alphasrv2100/alphasrv2100_abstract.html .